



## Article

# Ionospheric Kalman Filter Assimilation Based on Covariance Localization Technique

Jiandong Qiao <sup>1</sup>, Chen Zhou <sup>1,\*</sup> , Yi Liu <sup>1</sup>, Jiaqi Zhao <sup>2</sup> and Zhengyu Zhao <sup>2</sup><sup>1</sup> Department of Space Physics, School of Electronic Information, Wuhan University, Wuhan 430072, China<sup>2</sup> Institute of Space Science and Applied Technology, Harbin Institute of Technology, Shenzhen 518055, China

\* Correspondence: chenzhou@whu.edu.cn

**Abstract:** The data assimilation algorithm is a common algorithm in space weather research. Based on the GNSS data from the China Crustal Movement Observation Network (CMONOC) and the International Reference Ionospheric Model (IRI), a fast three-dimensional (3D) electron density nowcasting model for China and its adjacent regions was developed. Unlike the traditional Gaussian background covariance model, the error covariance of the IRI model, based on the IGS grid TEC data, is estimated in this work. Due to the large scale of the high-resolution covariance matrix, it cannot be stored and calculated directly on a personal computer. The covariance localization (CL) technique is introduced to sparse the covariance matrix while removing the pseudo-correlation of the covariance matrix. After localization, the covariance matrix can be converted into a sparse matrix for storage and calculation, which greatly reduces the computer memory requirement of the assimilation model and improves the calculation speed of the model. Based on this algorithm, a series of experiments were carried out in this work. The experimental results show that this algorithm can effectively assimilate the observed GNSS to the background field, make up for the temporal and spatial limitations of the observed data, and improve the accuracy of the ionospheric electron density nowcast. Compared with the digisonde observed foF2 (the critical frequency of the ionospheric F2 layer), the RMSE of the assimilation model is 0.44 MHz lower than that of the IRI model.

**Keywords:** data assimilation; electron density; covariance estimation; covariance localization



**Citation:** Qiao, J.; Zhou, C.; Liu, Y.; Zhao, J.; Zhao, Z. Ionospheric Kalman Filter Assimilation Based on Covariance Localization Technique. *Remote Sens.* **2022**, *14*, 4003. <https://doi.org/10.3390/rs14164003>

Academic Editors: Shuanggen Jin and Vladislav Demyanov

Received: 9 July 2022

Accepted: 11 August 2022

Published: 17 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Ionosphere Data Assimilation

The ionosphere is a highly variable environment that presents significant weather variations with latitude, longitude, altitude, local time, season, and solar and geomagnetic activities. In an effort to specify ionospheric weather, a large number of physics-based theoretical and numerical ionospheric models have been developed based on historical ionospheric data. Due to a lack of accurate estimation of ionospheric driving parameters and the use of average information of ionospheric-related parameters in the modeling process, although physics-based theoretical and numerical models can reproduce many climatic characteristics of the ionosphere, these models fail to reproduce the detailed changes of the ionospheric weather [1]. Data assimilation methods can assimilate continuous observations into the model to provide a global/regional ionospheric specification that is in good agreement with the model and observed data [2]. In recent years, the theory of ionospheric data assimilation has been rapidly developed. Using a data assimilation algorithm to build a model, which not only contains the internal physical process but also reflects the real observation, has become a growing interest in ionospheric research. Data assimilation methods, such as 3D-VAR [3], the Kalman filter [4–10] have been widely used.

The physics-based model is based on the dynamic equation. Given the external driver, it can describe the internal physical changes of the ionosphere and propagate the state forward in time. By assimilating single-source or multi-source ionospheric observation data

into the ionospheric physics-based model, the ion composition of the model is updated, which effectively improves the short-term prediction ability of the model [1,2,4,5,11]. In addition, some coupling models that combine different spatial domains (e.g., ionosphere–thermosphere, magnetosphere–ionosphere) have been developed to study the influence of the updated model state on the prediction results, with the ionospheric observation used as the boundary condition of the ionosphere part of the model [8,11–13]. Based on the thermosphere-ionosphere-electrodynamics general circulation model (TIEGCM), Chen et al. (2016) evaluated the impact of different assimilation-forecast cycle lengths on ionospheric predictions during geomagnetic storm conditions. Hsu et al. [9] evaluated the influence of the set's size on the assimilation effect by using the ensemble Kalman filter (EnKF). He et al. [10] first found that by using sparse matrix technology, the EnKF-based global ionospheric and thermosphere data assimilation system could be conducted without a supercomputer. Further, the prediction performance of this model was evaluated by He et al. [14].

The empirical model is a statistical model representing the average essence of observations, which fails to reproduce various short-term events. The empirical model does not include the physical process of ionospheric change; thus, it is unable to predict the ionosphere. Its prediction is usually realized by other algorithms, such as the Gaussian Markov algorithm [15]. However, due to its simple deployment and high computational efficiency, empirical models are also widely used in many fields. For example, the International Reference Ionosphere (IRI) model has been widely used in science, engineering, and education, and is generally superior to other empirical and theoretical models [16]. In recent event-based comparisons of the ionospheric models with ground and space measurements, within the framework of the CEDAR Electrodynamics Thermosphere Ionosphere (ETI) Challenge by the Community Coordinated Modeling Center (CCMC), the IRI was one of the two top performers [16]. Although the IRI model does not have a prediction ability, assimilating ionospheric observation data into the IRI model to improve the accuracy of the ionospheric nowcast has proved to be effective [6,16]. Based on the IRI model, a large number of global [17,18] or regional [7,19] three-dimensional ionospheric nowcasting models have been developed. This method can also be used as regional real-time TEC mapping technology [20]. Yue et al. [18] retrieved the occultation electron density profile by assimilating the occultation slant total ionospheric electron content (STEC) into the IRI model, and the results showed that this method is better than the traditional Abel integral inversion method.

### 1.2. Covariance Modeling

For the ionospheric nowcast based on the empirical model, the current challenges are the estimation of background-error covariance and the storage and calculation of a super-large-scale covariance matrix. In the traditional assimilation model, the Gaussian covariance estimation method is often used to estimate the background-error covariance matrix, where the Gaussian model is static and assumes that the ionospheric correlation length is fixed, and an exponential function is used to describe the variation in the ionospheric correlation coefficient with distance [2,3,6,7,16–19]. Generally, the ionosphere is anisotropic and changes with the longitude and latitude. The fixed correlation length cannot fully describe the change in the ionosphere. Forsythe et al. [21,22] estimated the correlation length of the IRI in horizontal and vertical directions based on the global ionosphere map (GIM) and incoherent scattering radar data. Their results showed that the correlation length of the ionosphere is also anisotropic, and the correlation region of the reference point should be an asymmetric ellipse. Wang et al. [23] generated the background dataset by disturbing the driving parameters of the IRI model, statistically obtained the covariance, and linearly combined it with the traditional Gaussian covariance to analyze the influence of the background model on the data assimilation results. However, the asymmetric elliptical correlation length model easily leads to the asymmetry of the covariance matrix. The

statistical covariance model has to calculate dozens to hundreds of sets in each iteration, which usually results in the time overhead being unacceptable in practical applications.

No matter which assimilation method is used, its core is the modeling of the covariance matrix. Although, many researchers use these similar algorithms to study the nowcast or prediction of regional or global ionospheric weather and have achieved good results. However, in these studies, based on the assumption of the Gaussian correlation of ionospheric error, the covariance is modeled by the Gaussian model with a fixed correlation length, and the full matrix is used for a direct calculation in the assimilation process. In a small area or low-resolution assimilation, the calculation time and resource consumption can be ignored. However, for large-scale or high-resolution assimilation systems, the dimension of the matrix rises sharply. Generally speaking, the number of elements of the covariance matrix is equal to the square of the number of variables. When the resolution in longitude and latitude is doubled, the covariance matrix is increased by 16 times, which is undoubtedly unacceptable for personal or small computers.

### 1.3. Innovation

In this paper, based on a large number of previous observation data, we measured the statistics of the errors of the IRI model under different solar activities, geomagnetic conditions, local time and other conditions, obtaining the correlation coefficient matrix under different conditions, and archived it. This correlation coefficient data file can be read and used according to the relevant conditions in the assimilation, which can avoid the errors introduced by the assumption of the ionospheric Gaussian correlation and fixed correlation length to a certain extent.

Covariance localization is a classical covariance filtering technology that can eliminate the long-distance pseudo-correlation of covariance and improve the performance of data assimilation. The localized covariance matrix is a matrix composed of a large number of 0 elements, which just meets the sparsity requirements of the sparse matrix. In previous studies, the correlation coefficient threshold is usually set to truncate the covariance matrix and then transform it into a sparse matrix, which may make the covariance matrix irreversible, resulting in filtering divergence. Truncating with the covariance localization function can ensure that the truncated covariance matrix is reversible and makes the assimilation system more stable.

All in all, the algorithm in this paper can take advantage of more ionospheric correlations and circumvent the use of a large-scale covariance matrix by applying the covariance localization and sparse matrix technology to optimize the assimilation algorithm, which would greatly reduce the computer memory requirement of the assimilation model and improve the calculation speed of the model. This provides a feasible method for regional GNSS ionospheric real-time monitoring.

## 2. Materials and Methods

### 2.1. Kalman Filtering Algorithm

The Kalman filtering algorithm is one of the common linear filtering and prediction methods. It is proposed by Kalman et al. [24]. It is the first sequential assimilation algorithm and is generally regarded as the theoretical basis of the sequential assimilation algorithm. In recent studies, the Kalman filter algorithm was widely used in data assimilation.

According to the Kalman filtering theory, the assimilation process of observed data can be expressed as:

$$X_t^a = X_t^b + K(Y_t - HX_t^b), \quad (1)$$

$$K_t = B_t H^T (H B_t H^T + R)^{-1}, \quad (2)$$

$$B_t^a = B_t - B_t H^T (H B_t H^T + R)^{-1} H B_t, \quad (3)$$

$$X_{t+1}^b = M X_t^a + w_t, \quad (4)$$

$$B_{t+1}^b = MB_{t+1}M^T + Q, \quad (5)$$

where  $X^a$  is the electron density after assimilation,  $X^b$  is the electron density of the background field given by the IRI model,  $B$  is the covariance matrix of the error of the background field,  $Y$  is the vertical total electron content (VTEC) of the GNSS observation,  $H$  is the observation operator, and  $R$  is the covariance matrix of the error of the observation field.  $K$  is the gain matrix, which plays a role in adjusting the background field from the observed data.  $B_a$  is the analysis field covariance matrix after assimilation.  $M$  is the state transition matrix, and  $w$  is the state transition error.

In data assimilation, the result of data assimilation can be considered as the weighted average of the observation and background. Therefore, whether the error of the background field and observation field can be accurately described significantly affects the results of the assimilation experiment. In previous studies, the errors among the observed data are considered to be uncorrelated and unbiased, and the observation error is considered proportional to the square of the observation value [6,15,19,25,26]. Therefore, the observation-field error covariance matrix  $R$  is the diagonal matrix, and the element of the diagonal is proportional to the observation square.  $R$  is expressed by Equation (6):

$$R_{ij} = \sigma_i \sigma_j \rho_{ij} = \begin{cases} \eta_o Y_i Y_j, & (\rho_{ij} = 1, i = j) \\ 0, & (\rho_{ij} = 0, i \neq j) \end{cases} \quad (6)$$

where  $i$  and  $j$  represent the observation point,  $R_{ij}$  is the error covariance among the observational points,  $\rho_{ij}$  is the correlation coefficient, and  $\eta_o$  is a proportionality coefficient.

Since the errors of the observation data are not correlated with each other, only the covariance matrix of background field errors can be used to transfer the spatial correlation information to the assimilation results. The background error covariance model is based on Gaussian theory (GBEC) and statistics (SBEC), two common modeling methods in ionospheric data assimilation. The SBEC is a common modeling method of the EnKF assimilation algorithm. It needs to calculate multiple groups of background data as samples and then obtain the background-error covariance matrix through statistical analysis. The GBEC is a common method to model the background field error covariance matrix in the Kalman filter assimilation algorithm. The SBEC is more flow-dependent with enough samples, which can obtain a more accurate covariance matrix. However, in each assimilation cycle, the SBEC needs to calculate dozens to hundreds of groups of background data samples to estimate the background-error covariance matrix. For large areas or high-resolution, it needs a lot of calculation time. The GBEC only needs to calculate a group of background data and then estimate the background-error covariance matrix through empirical formulas, which can obviously reduce the calculation time. The Gaussian error covariance model assumes a Gaussian correlation between background data, and the covariance is proportional to the square of the observed value. The calculation method is as follows:

$$B_{ij} = \sigma_i \sigma_j \rho_{ij}, \quad (7)$$

$$\sigma_i^2 = \eta_b X_i^b X_i^b, \quad (8)$$

$$\rho_{ij} = e^{-\Phi_{ij}^2/2L_\Phi^2} e^{-\theta_{ij}^2/2L_\theta^2} e^{-H_{ij}^2/2H_\theta^2}, \quad (9)$$

where  $B_{ij}$  is the error covariance between background data,  $\rho_{ij}$  is the correlation coefficient,  $\eta_b$  is the proportional coefficient,  $\Phi_{ij}$ ,  $\theta_{ij}$  and  $H_{ij}$  are the distance between the grid points in the longitude, latitude and height directions. We assume that the ionospheric correlation is separable in the horizontal and vertical directions, so we calculate the ionospheric correlation in the horizontal and vertical directions, respectively. Because it is difficult to obtain the vertical structure information of the ionosphere by conventional observations, the vertical structure of the ionosphere can only be observed by a few devices, such as the digisonde or incoherent scattering radars. Therefore, for the vertical direction, we refer to

the previous research experience and the actual assimilation effect and take the vertical correlation length as 40 km.

In previous studies, it was generally considered that the ionosphere was isotropic, or isotropic toward the direction of longitude, latitude and altitude, and the correlation length was fixed and did not change with time, latitude and longitude, which did not conform to the actual characteristics of the ionosphere. In this work, the CODE grid TEC data, provided by IGS from 2003 to 2018, was used as the truth values to evaluate the errors of the IRI-2016 model. The dataset was divided according to the solar flux, geomagnetic AP index, season and local time, and the ionospheric correlation coefficients under different conditions were statistically analyzed, and then a data file was formed. Each bin has about a few hundred samples. This data file can be read and used according to the relevant conditions in the assimilation. The dataset division conditions are shown in Table 1.

**Table 1.** Data division conditions.

Season			
Spring (March to May)	Summer (June to August)	Autumn (September to November)	Winter (December to February)
F10.7			
Low (0–100)	Moderate (100–150)	High (>150)	
AP index			
Low (0–30)	Moderate (30–50)	High (>50)	
Time			
0–24 h/2 h			

The latitude and longitude resolutions of the assimilation model in this work are  $1^\circ \times 1^\circ \times 30$  min. In order to obtain the correlation coefficient matrix, we first matched the IGS grid and assimilation grid into space and time. Since the resolution of the IGS data is  $5^\circ \times 2.5^\circ \times 2$  h, we performed linear interpolation on the grid data of IGS in the latitude and longitude directions to change the resolution to  $1^\circ \times 1^\circ \times 30$  min. When estimating the covariance, the usual method is the empirical covariance estimate, which is the maximum likelihood estimator of the covariance and is unbiased, i.e., it converges to the true covariance when given many observations. However, when the number of the sample is smaller than that of the variable, the empirical covariance estimation is not the optimal estimate of the covariance, which may cause the covariance matrix to be less than the rank, resulting in the matrix being irreversible, and regularization to reduce its variance is usually beneficial. In this work, the Ledoit–Wolf shrunk covariance estimator was used for estimating the error covariance of the IRI model [27]. The Ledoit–Wolf shrunk covariance estimation is a generalized linear combination based on the covariance matrix and identity matrix. When the sample size is smaller than the variable dimension, the performance is better, and the covariance matrix can be guaranteed to be a positive-definite matrix. The calculation method is as follows:

$$dN_i(t) = N_{IRI}(t) - N_{IGS}, \quad (10)$$

$$D_t(t) = dN_i(t) - \overline{dN_i}, \quad (11)$$

$$M = D * (D)^T / (M - 1), \quad (12)$$

$$M = (1 - sk) \times M + sk \times \mu \times eye(n), \quad (13)$$

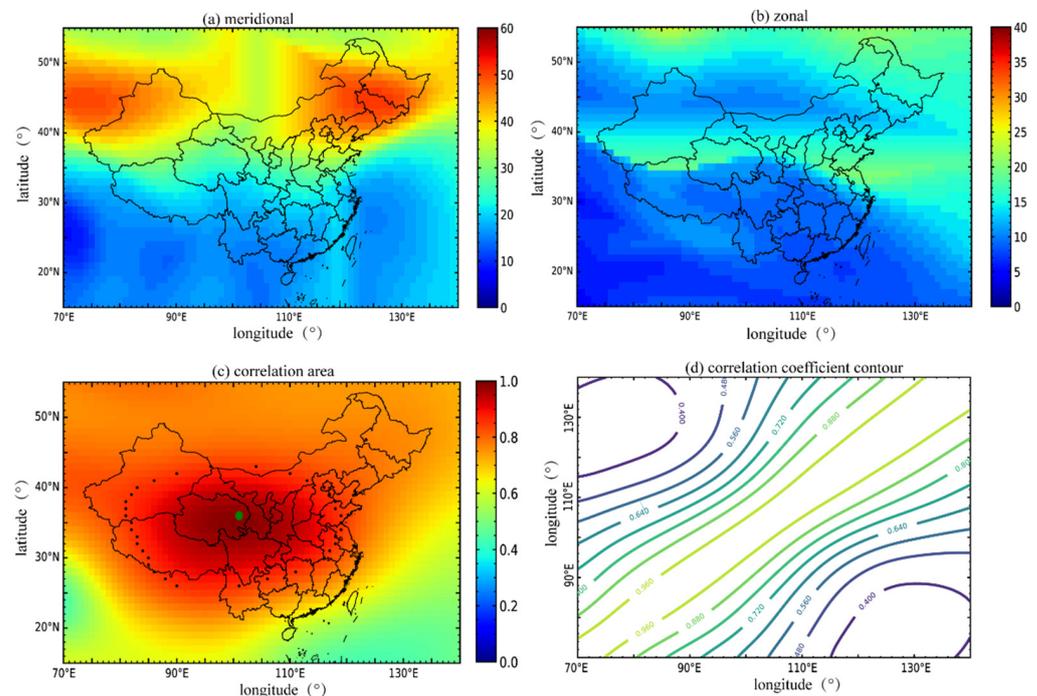
$$\mu = trace(M) / n, \quad (14)$$

where,  $N_{IRI}$  is the IRI sample set,  $N_{IGS}$  is the IGS sample set,  $sk$  is the shrinkage coefficient [27],  $eye$  is the identity matrix, and  $trace$  is a trace of the matrix.

Therefore, Equation (14) can be redefined as:

$$\rho = M / (\sqrt{\text{trace}(M)^T * \text{trace}(M)}), \quad (15)$$

Figure 1 shows the distribution of the correlation coefficient and correlation length under the magnetic static period ( $AP < 30$ ) and middle solar activity ( $100 < F10.7 < 150$ ) condition during winter (January–March) at 00:00 UT (universal time). Figure 1a is the maximum meridian (the azimuth angle is  $90^\circ$  and  $270^\circ$ ) correlation length distribution; Figure 1b is the distribution of the maximum zonal (the azimuth is  $0^\circ$  and  $90^\circ$ ) correlation length; Figure 1c is the correlation coefficient distribution between the reference point and its surrounding grid points, and the black dot line is the dividing line with a correlation coefficient of 0.8. Figure 1d is the contour diagram of the meridian correlation coefficient at  $70^\circ\text{E}$  and  $15^\circ\text{N}$ . It can be seen from Figure 1 that the ionosphere is anisotropic; as a result, the ionospheric correlation coefficient models under different conditions can more accurately describe the ionization changes and thus improve the effects of the assimilation.



**Figure 1.** The distribution of the correlation coefficient and correlation length under magnetic static period ( $AP < 30$ ), and middle solar activity ( $100 < F10.7 < 150$ ) condition during winter (January–March) at 00:00 UT. Where (a) is the distribution map of meridional correlation length; (b) is the zonal correlation length distribution map; (c) is the reference points and their related regional distributions; (d) is the meridian contour distribution at  $70^\circ\text{E}$  and  $15^\circ\text{N}$ .

It is generally believed that there is a strong correlation between the ionospheric state change and time. In this work, the Gaussian Markov algorithm was used to describe the correlation. In the hidden Markov model, the probability distribution of the next state can only be determined by the current state. Therefore, the time-decay error term is added to the calculation result of the background field model as the prediction equation of the Kalman filter. The error term is the difference between the state analysis value of the current time step and the model-calculated value. The attenuation coefficient is an exponential function related to the time scale, and 3 h (hour) in the geomagnetic calm period and 1 h

in the geomagnetic storm period are selected [19]. Therefore, Equations (4) and (5) can be rewritten as Equations (16) and (17):

$$X_{t+1}^b = (x_t^a - x_t^b) * e^{-\frac{\Delta t}{\tau}} + x_{t+1}^b, \quad (16)$$

$$B_{t+1}^b = B_t^a * e^{-\frac{2\Delta t}{\tau}} - B_t^b * e^{-\frac{2\Delta t}{\tau}} + B_t^{b+1}, \quad (17)$$

where  $X_{t+1}^b$  is the background TEC, and  $\tau$  is the ionospheric correlation time varying with local time, and represents the influence of the assimilation results of the current step on the background state of the next step. The value in this paper is 5 h [1]. We introduced relevant scale factors that vary with local time. The formula is as follows:

$$flt = 1/4 * \cos(\pi/12 * (t - 14)) + 3/4, \quad (18)$$

Because the observation sites are sparse, the observation field and the background field usually have different dimensions. Therefore, the observation operator H is also an important parameter, which can map the background field into the observation field. Generally, H is a matrix with a different number of rows and columns. The number of rows is equal to the number of observed grid points, and the number of columns is equal to the number of grid points in the background field. When the observed value does not exist, the corresponding column number element is set to 0; thus, H is a sparse matrix. In this work, the state quantity of the observed field is VTEC, and the state quantity of the background field is the electron density. Therefore, H represents the integral from the electron density to the total electron content (TEC), and the non-zero element in H is the length of the grid point.

## 2.2. Covariance Localization (CL)

Covariance models always introduce pseudo-correlation when estimating the background-error covariance matrix. That is, there is a pseudo-correlation between the state variables that are physically unrelated or spatially distant. As the assimilation process goes on, the background-error covariance is underestimated by the system, the weight of observation data becomes smaller and smaller, and finally, it is completely ignored, leading to the divergence of filtering. In order to reduce the influence of the pseudo-correlation on the assimilation process and improve the assimilation results, two methods, covariance inflation [28] and localization [29], are proposed. Covariance inflation is a method used to correct the underestimation of the background-error covariance matrix, which is mainly used in the ensemble Kalman filter. The principle is that the dispersion of each ensemble member from the mean is multiplied by an inflation factor slightly greater than 1 to solve the large sampling error caused by the limited sample size. At present, two localization methods: covariance localization (CL) [30] and local analysis (LA) [31], are widely used.

CL is a covariance filtering method, which uses localized radius to truncate the distance pseudo-correlation in the background-error covariance matrix and improves the quality of the background-error covariance matrix, making the covariance matrix more sparse. CL is typically carried out by applying a Schur (Hadamard) product between a correlation matrix  $\rho$  with distance-decreasing entries and an ensemble-estimated covariance matrix.  $\rho$  is a Gaspari–Cohn (GC) five-order piecewise polynomial function, and the calculation method is as follows [32]:

$$B_t = \rho \circ B_t, \quad (19)$$

$$\rho = \begin{cases} -\frac{1}{4} (||z||/c)^5 + \frac{1}{2} (||z||/c)^4 + \frac{5}{8} (||z||/c)^3 - \frac{5}{3} (||z||/c)^2 + 1, & 0 \leq ||z|| < c \\ \frac{1}{12} (||z||/c)^5 - \frac{1}{2} (||z||/c)^4 + \frac{5}{8} (||z||/c)^3 + \frac{5}{3} (||z||/c)^2 - 5 (||z||/c) + 4 - \frac{2}{3} (c/||z||), & c \leq ||z|| < 2c \\ 0, & 2c \leq ||z|| \end{cases}, \quad (20)$$

where  $c$  is the localization radius, the value is 15; that is, the points whose horizontal distance is greater than are filtered out.  $\rho$  from  $\rho(0) = 1$  smooth descent to  $\rho(2c) = 0$ . The background-error covariance matrix  $B$  is a non-negative-definite matrix and  $\rho$  is a positive-definite matrix, so the Schuur product of  $B$  and  $\rho$  is also a positive-definite matrix. Thus, Equation (2) can be rewritten as:

$$K_t = (\rho \circ B_t) H^T (H(\rho \circ B_t) H^T + R)^{-1}, \quad (21)$$

Figure 2a,b presents the distribution of the meridional correlation coefficients obtained from the statistical results and are localized by covariance. It can be seen from Figure 2a that the meridional correlation tends toward 0 with the increase in the distance. The local correlation coefficient is close to the original one in the central part, however, the attenuation is faster. The localization radius can control the size of the truncated region. The larger the localization radius is, the larger the correlation region will be, and the smoother the result after assimilation will be. However, the increase in the localization radius will lead to a sharp increase in the size of the covariance matrix. Therefore, it is important to choose a reasonable localization reification radius in practical applications. Figure 2c shows the correlation coefficient curves of the longitude before and after localization. Figure 2d shows the correlation coefficient curves of zonal and meridional. Figure 2e is the change curve of the correlation coefficient before and after localization at  $5^\circ$  from the reference point. Figure 2f shows the change curve of the correlation coefficient in different seasons at  $5^\circ$  from the reference point. The curve after localization is located below the before, but this does not exactly mean that the distribution fully depends on the localization radius. It can be seen from Figure 2c–f that the localized correlation coefficients are different for different directions, times and seasons. This also means that our covariance model contains ionospheric anisotropy information. When we estimate the correlation coefficient under different conditions, the sample size is much smaller than the variable dimension. The localization function can be regarded as a tapering covariance regularization method, which can provide some prior information for sample covariance [33]. This can not only reduce the error problem when estimating the covariance from small samples, but it can also avoid a series of abnormal problems caused by threshold truncation; for example, the covariance matrix is not positive-definite. However, whether the local function we use is optimal or whether we can estimate the optimal local function needs to be evaluated in the follow-up work.

### 2.3. Sparse Matrix Compression Algorithm

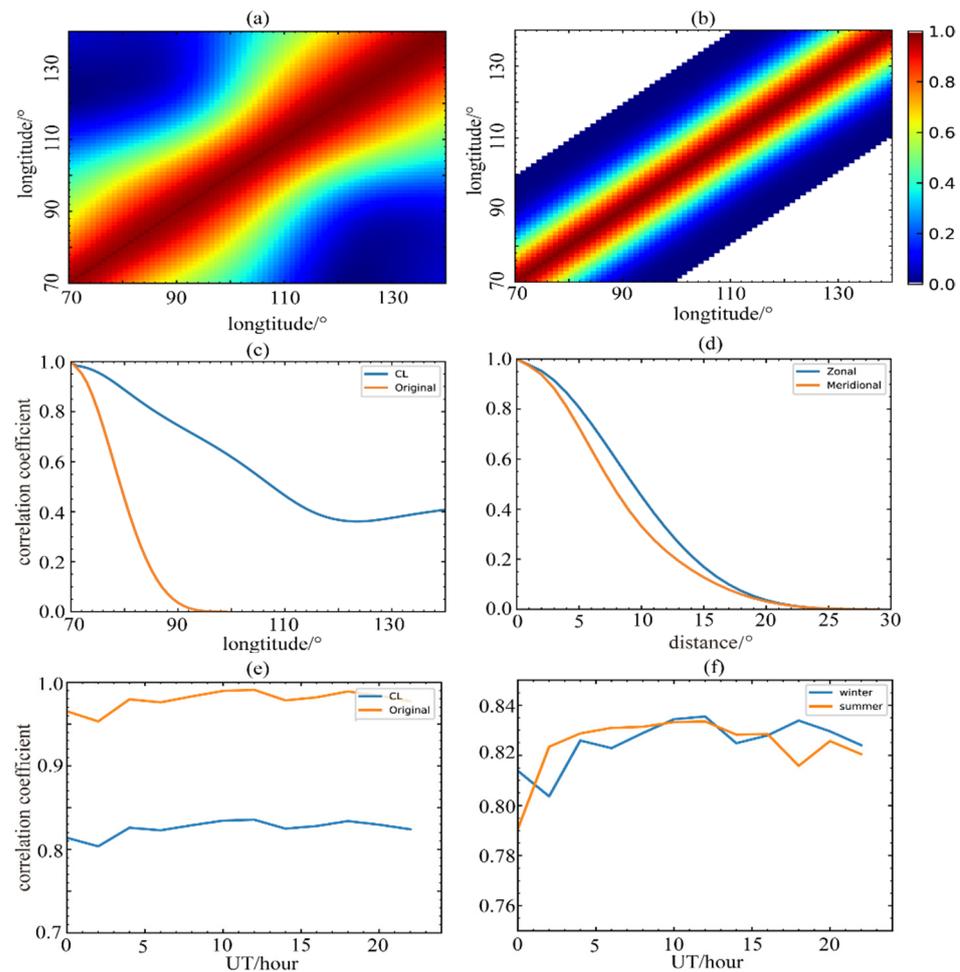
In this paper, the longitude and latitude resolution is  $1^\circ \times 1^\circ$ , the altitude resolution is 10 km and the time resolution is 30 min. The dimension of the background field data is  $41 \times 71 \times 45$ , and the size of the covariance matrix is  $130,995 \times 130,995$ . When calculating and storing with float64 type, the size of the matrix is about 128 G. In the matrix operation, the whole matrix is needed to be read into the memory first, and then the operation is carried out, which cannot be realized by a personal computer, and the time-cost of complex matrix operations, such as the inverse of the whole matrix is very large.

A sparse matrix is a matrix consisting of almost zero values. A sparse matrix differs from most non-zero valued matrices, called dense matrices. When storing a sparse matrix, memory must be allocated for each 32-bit or even 64-bit zero value in the matrix, which is a waste of memory resources, since these zero values do not contain any information. We can use compression technology to minimize the amount of data needed to be stored. Almost all algorithms require the data matrix to exist in memory beforehand, which means that when the data matrix cannot be fully stored in memory, the calculation process will be interrupted. One advantage of converting from dense to sparse matrices is that sparse matrices can be compressed into a suitable memory size in most cases. The sparsity of the

matrix can be quantified by a score, which is the number of non-zero values in the matrix divided by the total number of elements in the matrix.

$$\text{sparseness} = \frac{n_0}{n}, \quad (22)$$

where  $n_0$  is the number of non-zero elements in the matrix, and  $n$  is the total number of elements in the matrix.



**Figure 2.** (a) The distribution of meridional correlation coefficients at 70°E, 15°N obtained from statistics; (b) the distribution of meridional correlation coefficients at 70°E, 15°N, localized by covariance; (c) the correlation coefficient curves of longitude before and after localization; (d) the correlation coefficient curves of zonal and meridional; (e) the change curves of the correlation coefficient before and after localization at 5° from the reference point; (f) the change curves of the correlation coefficient in different seasons at 5° from the reference point.

A solution for representing and processing a sparse matrix is to use a data structure that ignores zero values and stores only the non-zero values in the sparse matrix to represent the sparse matrix. Commonly used sparse matrix compression formats include the coordinate format (COO), compressed sparse row format (CSR), compressed sparse column format (CSC), and the list of lists format (LIL). The CSR sparse matrix uses three one-dimensional arrays to represent non-zero values, row ranges, and indices. This format requires the matrix elements to be stored in row order and the elements in each row to be stored out of order. For each row, a pointer is used to indicate the starting position of the element. The CSR format supports an efficient arithmetic operation, efficient row slicing and fast matrix operation. Therefore, storing the sparse matrix in CSR format can significantly improve

the calculation speed of the assimilation model; however, the time-cost of changing the sparsity structure of the CSR matrix is very expensive. For a super-large covariance matrix, it cannot be transformed directly, and elements need to be added one by one through a loop. Thus, by saving directly in CSR format, the matrix time-cost is very expensive. Therefore, it is necessary to construct the sparse matrix of LIL formats through a circular method before converting them to CSR formats. The LIL format uses two nested lists to store the sparse matrix, which means that data hold the values of the non-zero elements in each row, and the rows hold the column numbers of the non-zero elements in each row (the column numbers are sorted sequentially). This format is good for adding elements one by one and obtaining row-specific data quickly, but the LIL matrix multiplication is extremely inefficient. The problem of storage and computing speed regarding a super-large sparse matrix can be effectively solved by first constructing the LIL-format sparse matrix through the cyclic method and then converting it into the CSR-format sparse matrix for calculation. In this paper, the sparse matrix is firstly constructed in LIL format and then converted into the CSR format for calculation.

### 3. Data

#### 3.1. Background Data

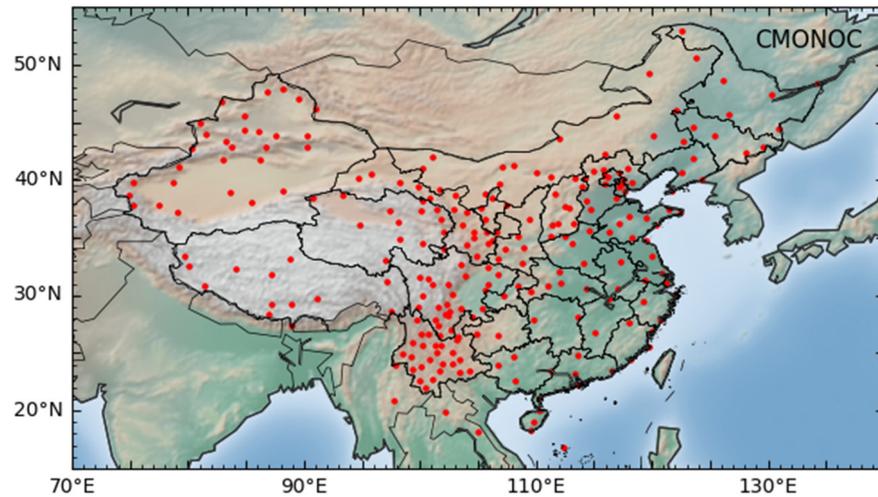
The selection of background model and observation data directly determines the quality of assimilation-forecast results. The IRI model is a commonly used ionospheric model in ionospheric research, and it integrates a variety of ionospheric empirical models with good performance, such as the nequick model. Therefore, the IRI [34,35] model is chosen as the background model for the assimilation in this work. Since this work focuses on the ionospheric changes in China and its surrounding areas, the latitude and longitude ranges of the background field data are 15°N–55°N and 70°E–140°E, with a step size of 1°, and a height range of 60–500 km, with a step size of 10 km. We assume that the electron density above 1000 km is negligible, and the electron content in the IRI model above 500 km is proportional to the observation. Therefore, this part of electron content was subtracted from the GNSS observations and then assimilated into the model. Therefore, the size of background-error covariance matrix B is  $130,995 \times 130,995$ . The space division is shown in Table 2.

**Table 2.** Spatial and temporal division of electron density.

Parameter	Longitude	Latitude	Height	Time
range	70°E–140°E	15°N–55°N	60–500 km	0–24 h
step	1°	1°	10 km	0.5 h

#### 3.2. Observation Data

Observational data from 261 GNSS observatories at CMONOC were used in this work. The sampling interval of the GNSS receiver is 30 s. The distribution of the GNSS stations in CMONOC is shown in Figure 3.



**Figure 3.** Geographical distribution of CMONOC GNSS stations.

The GPS dual-frequency receiver can receive L1 (1575.42 MHz) and L2 (1227.6 MHz) carrier signals simultaneously. By solving the group delay and phase delay data of the GNSS dual-frequency signals jointly, the electron density integral along the propagation path of each satellite signal, i.e., the slant TEC (STEC), is obtained. Based on the least square principle, the carrier-phase observation is used to smooth the pseudo-range observation, and the high-precision absolute TEC data can be obtained.

$$STEC_L = \left[ \left( \frac{f_2^2}{f_1^2 - f_2^2} \right) \frac{2f_1^2}{K} \right] (P_2 - P_1), \quad (23)$$

$$STEC_P = \left[ \left( \frac{f_2^2}{f_1^2 - f_2^2} \right) \frac{2f_1^2}{K} \right] (L_1\lambda_1 - L_2\lambda_2), \quad (24)$$

where  $STEC_L$  is the integral of the electron density along the propagation path obtained by the differential pseudo-distance observation, known as absolute TEC;  $STEC_P$  is the product component of the electron density along the propagation path obtained by the differential carrier-phase detection, known as relative TEC;  $P_1$  and  $P_2$  are the pseudo-distances measured by the dual-frequency GNSS receiver;  $\lambda_1$  and  $\lambda_2$  are the wavelengths of the carrier;  $L_1$  and  $L_2$  are carrier phases;  $K$  is a constant related to the plasma frequency and electron concentration, generally,  $K$  is  $80.62 \text{ m}^3\text{s}^{-2}$ .

VTEC is the current common GNSS level two or three product data. For many applications, including shortwave communications, radar frequency selection, refraction error correction, etc., VTEC cannot be used directly. Therefore, instead of using STEC directly for assimilation, we used VTEC, which provides a feasible method to make full use of this part of the data. STEC can be converted into a vertical TEC (VTEC) by Equation (25). In the process of solving the vertical TEC, the ionospheric-charged particles are generally considered to be concentrated in a thin shell that is concentric with the Earth at an altitude of 350–450 km, which is set as 400 km in this work. According to this model, we can convert the STEC to TEC on the ionospheric thin-layer model with the same plane coordinates as the ionospheric penetration point (IPP), i.e., the vertical TEC (VTEC) at the pierce point position.

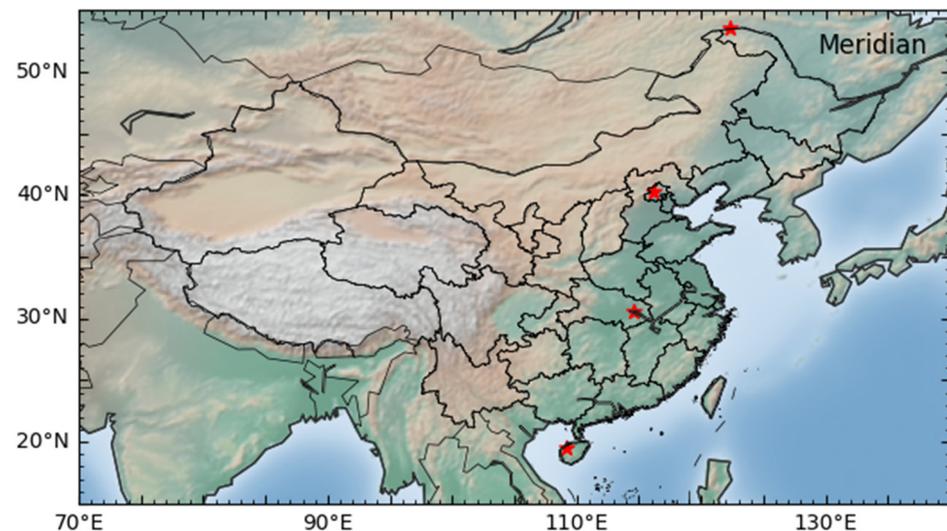
$$VTEC = (STEC - B_S - B_R) \sqrt{1 - \left( \frac{R_e \cos(E_i)}{R_e + h_s} \right)^2}, \quad (25)$$

where  $R_e$  is the radius of the Earth and  $h_s$  is the height of the puncture point.  $E_i$  is the observation elevation,  $B_S$  is the hardware error of the GNSS satellite transmitter, and  $B_R$  is the hardware error of the ground receiver. The hardware error is estimated by the fifth-order spherical harmonic function model, and the estimation strategy is a piecewise constant

method (PWC). Since only TEC is concerned in this paper, the combined estimation of the satellite and receiver errors is used to obtain the total hardware error  $B_S + B_R$ .

### 3.3. Digisonde Data

The ionospheric digisonde can measure the time delay of the ionospheric echo reaching the receiver by transmitting a sweep-frequency pulse vertically from the ground and obtaining the information of the ionospheric virtual height, varying with the frequency at each frequency point. Using the embedded software from Digisonde, the ionospheric characteristic parameters and electron density profile can be measured and calibrated automatically, and the SAO data of digisonde can be generated. In this work, the SAO data of ionospheric digisonde were used as validation data, which was obtained from the Chinese Meridian Project [23]. The SAO file records the ionospheric characteristic scalar, which can be used to obtain information, such as the height, electron density and critical frequency of the ionosphere. The distribution of the digisonde stations is shown in Figure 4.



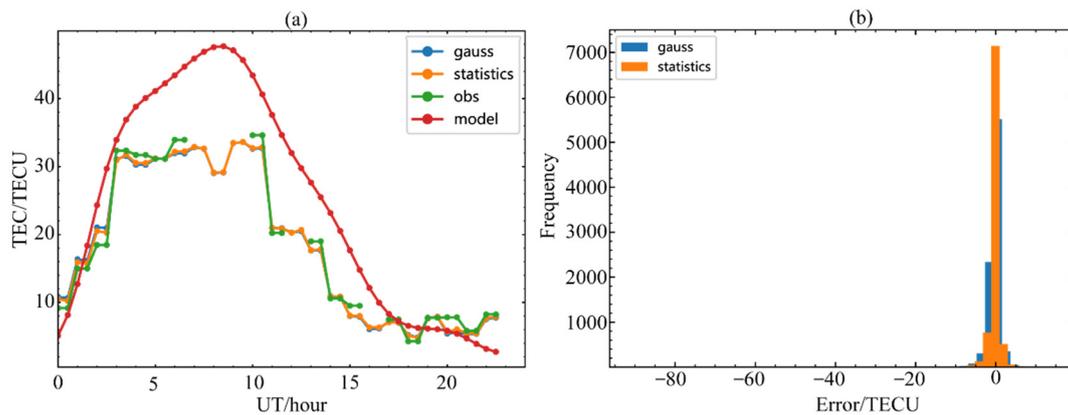
**Figure 4.** Geographical distribution of digisonde stations.

## 4. Results

In this work, a complete 3D electron density nowcast model for China and its adjacent regions was constructed based on the CL Kalman filter and sparse matrix technique. The background field was calculated by using the IRI model, which can be used to calculate the complete three-dimensional regional distribution of electron density. The original observation data were provided by CMONOC. Different from the background field data, GNSS cannot observe the vertical structure of the ionosphere, but it can obtain the ionospheric delay by phase- and pseudo-distance and then obtain the VTEC of the ionosphere. The observation stations have obvious discreteness and uneven distribution, so the observation field is a discrete incomplete VTEC map. In order to verify the effect of the assimilation inversion, the VTEC maps and electron density profiles of China and its adjacent regions are illustrated in this work. In the first part, we compared the performance of our covariance modeling method with that of traditional modeling methods. In the second part, we integrated the electron density of the background field and the assimilated electron density in the vertical direction, respectively, to obtain the background-field VTEC map and the assimilated VTEC map and compare them with the observation VTEC map. In the last part, the assimilation results were compared with the digisonde data of Zuoling in Wuhan, Changping in Beijing, Mohe in Heilongjiang and Fuke in Hainan. The three parts are described as follows:

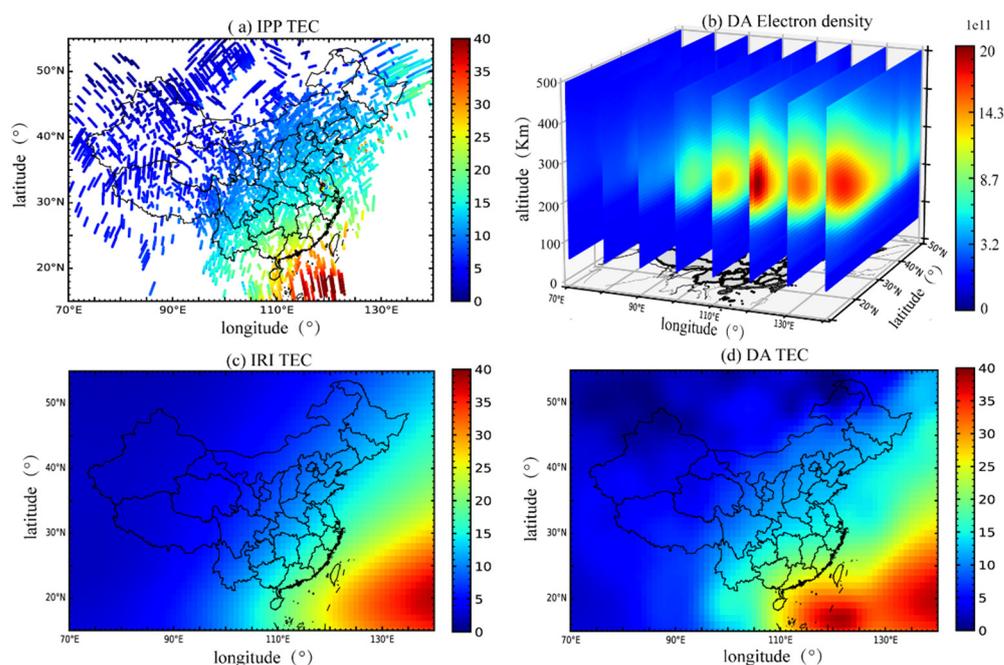
(1) We only modeled the ionospheric horizontal correlation coefficient under different conditions. Due to the lack of relevant observation data, the vertical direction correlation

refers to the previous research experience. Therefore, in order to discuss the difference, we ignored the height direction and compared the TEC results after the assimilation caused by the different covariance modeling methods. Figure 5a is the comparison curve of TEC for the different covariance modeling methods, which were observed and modeled for one day at 25°N, 90°E. The red curve is the simulation result of the model, the orange curve is the calculation result of statistical covariance modeling, the blue is the calculation result of the Gaussian model, and the green is the observation. Figure 5b is the TEC error statistical results of the different modeling methods in one day. It can be seen that our model will be improved compared with the Gaussian model in most cases. The absolute average error is smaller, and the error is reduced by 0.3 TECU.



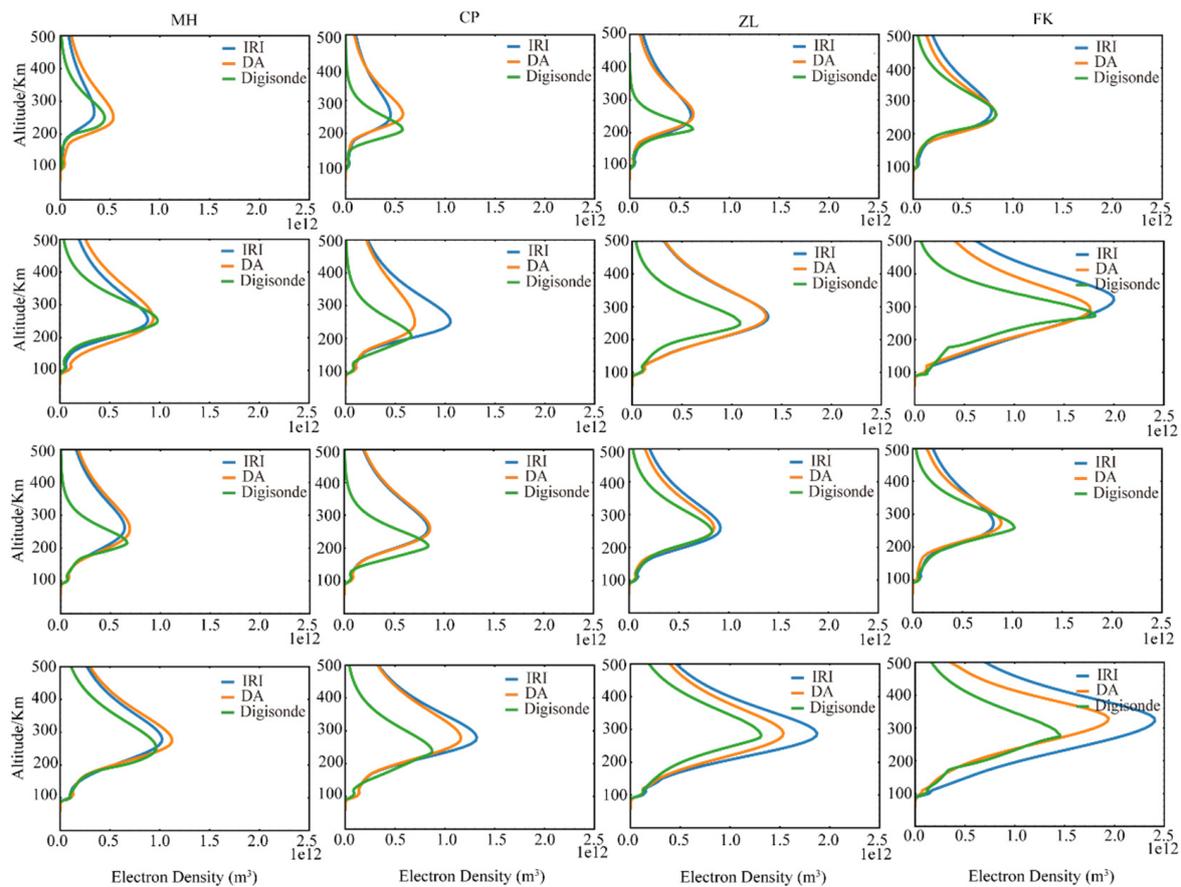
**Figure 5.** The comparison of TEC by different covariance modeling methods. (a) The comparison curve of TEC for different covariance modeling methods, observed and modeled in one day at 25°N, 90°E; (b) The TEC error statistical results of different modeling methods in one day.

(2) We pre-processed the observation data before assimilation. In the time window (30 min), a large number of data fall on the same grid point. For all observations in the same grid, we eliminated the data with a large deviation from the average value and then adopted the average value of the remaining observations as the value of the grid. Because the observation and background fields have different dimensions, the assimilation result was consistent with the background field dimension. In order to compare with the observation field, the 3D background field data and assimilation results were reduced in dimension. That is, the vertical integration was carried out to obtain the 2D VTEC distribution of China and its adjacent regions. Figure 6 shows the comparison of the assimilation results at 00:00 on 1 January 2015: (a) is the IPP TEC distribution by GNSS; (b) is the electron density map; (c) is the IRI TEC map; (d) is the TEC map after assimilation. We can see from Figure 6 that the TEC after assimilation is more detailed than the TEC calculated by the IRI. As shown in Figure 6a, the TEC is abnormal within the longitude and latitude zones of 15°–20°N and 110°–120°E and the total electron content increases sharply in this region, which cannot be simulated by the IRI model. However, the TEC performance in this region can be clearly seen in the TEC map after assimilation, as presented in Figure 6d. Through data assimilation, TEC anomalies in this abnormal region are assimilated into the electron density distribution. As shown in Figure 6b, the maximum electron density appeared near 120°E. This indicates that data assimilation can effectively assimilate 2D observation data into a 3D background model to improve the accuracy of the background model. However, due to the lack of GNSS observation stations in low and high latitudes of China, the electron density distribution in these regions can only be corrected by nearby observations. When the distance is greater than the localization radius, the two points in space will not be correlated, and the observation has only weak or no effect on the model. This means that the spatial and temporal distribution of the electron density in these regions can only be obtained through the background model.



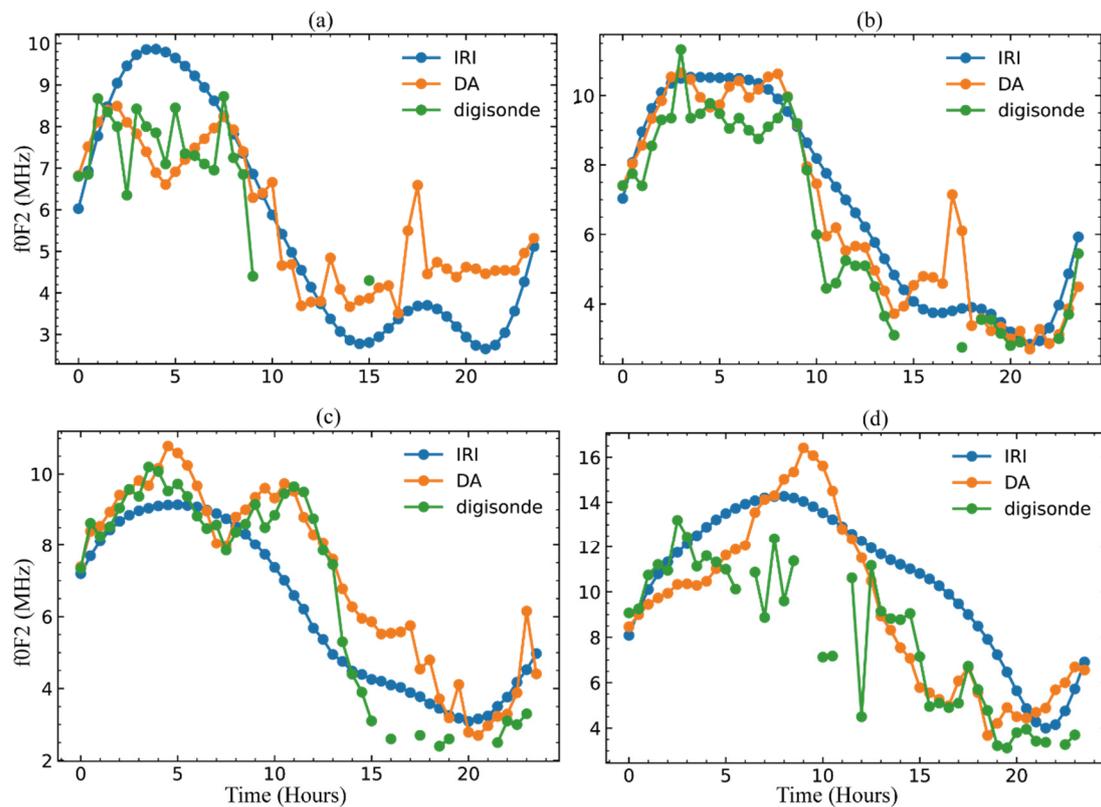
**Figure 6.** Electron density assimilation results at 00:00 UT on 1 January 2015. (a) VTEC of pierce point; (b) electron density distribution after assimilation (upper right); (c) VTEC calculated by IRI model (lower left); (d) VTEC after assimilation (lower right).

(3) Although the VTEC maps can describe the horizontal distribution of ionospheric electrons, they cannot accurately describe the internal vertical structure of the ionosphere. The electron density profile can present the internal structure of the ionosphere effectively and show the peak concentration and height information of the electron density in the E, F1 and F2 layers. Additionally, it can indirectly give the critical frequency of the ionosphere, foF2. In Figure 7, the first and second rows are the electron-density profile contrast diagrams at 00:00 UT and 06:00 UT on 1 January 2015, respectively. The third and fourth rows show the electron-density profile contrast diagrams at 00:00 UT and 06:00 UT on 17 March 2015, respectively. The assimilation and IRI results are compared with the digisonde observations at Changping (CP) in Beijing, Mohe (MH) in Heilongjiang, and Zuoling (ZL) in Wuhan and Fuke (FK) in Hainan. The green curve is the observational result of digisondes, the blue curve is the IRI modeling results, and the orange curve shows the assimilation results. As shown in Figure 7, the difference between the peak value of electron density after assimilation and the digisonde observation is smaller in most cases, which indicates that the algorithm can effectively assimilate the observation data into the background field. However, the peak heights of the electron density profile after assimilation are still consistent with the IRI model, which indicates that the peak height accuracy of electron density cannot be improved effectively. This is because the GNSS observation can only obtain TEC (the electron density integral), while it cannot supply the vertical structure of the ionosphere. Thus, the assimilation model depends on the ionospheric vertical structure information from the background model. In addition, the assimilation results are closer to the digisonde measurements below the F2 peak electron density, while large differences are found above the F2 peak electron density. This may be because the digisonde can only detect the structure below the F2 peak ionospheric electron density but cannot directly obtain the information above the F2 peak value. The profile above the F2 peak of the digisonde is obtained by software extrapolation, which may introduce errors. Moreover, the TEC observed by GNSS is from the ground to the satellite altitude, while the ionospheric model can only calculate several thousand kilometers. This leads to the model overestimating the electron density above F2.



**Figure 7.** Comparison of ionospheric electron density profiles. The green curve is the observational result of digisondes, the blue curve is the IRI modeling results, and the orange curve shows the assimilation results. The first and second rows are the electron-density profile contrast diagram at 00:00 UT and 06:00 UT on 1 January 2015, respectively. The third and fourth rows show the electron-density profile contrast diagram at 00:00 UT and 06:00 UT on 17 March 2015, respectively.

The foF2 is an important parameter in the ionosphere, and it is related to the peak electron density. In order to illustrate the influence of the assimilation model on peak electron density, Figure 8 shows the comparison of the foF2 variation at CP (a) and ZL (b) on 1 January 2015, and at MH (c) and FK (d) on 17 March 2015. Compared with the digisonde observations at these four stations, the RMSE (root mean square error) of the IRI model is 1.63 MHz, 1.28 MHz, 1.33 MHz, and 3.25 MHz, while the RMSE of the assimilation results is 0.84 MHz, 0.98 MHz, 1.12 MHz, 2.76 MHz, respectively, whose average RMSE is decreased by 0.44MHz. As presented in Figure 8, foF2, after assimilation, is closer to the digisonde observations. The changing trend of the IRI modeling results is consistent with the digisonde observations; however, it cannot accurately describe the changed details of the foF2. This result demonstrates that the assimilation model can effectively improve the accuracy of the electron density peak of the background model and can obtain more details.



**Figure 8.** Comparison of foF2 variation at CP (a) and ZL (b) on 1 January 2015 and at MH (c) and FK (d) on 17 March 2015.

## 5. Conclusions

In the present work, we analyzed the feasibility of the Kalman data assimilation algorithm to reconstruct the ionospheric electron density distribution in China and its adjacent areas. Based on the 15 years of IGS data from 2003 to 2018, the errors of the IRI model in different solar and geomagnetic conditions, seasons and local time were estimated, and the correlation coefficient model, adapted to different conditions, was formed. Using the correlation coefficient model and 261 GPS stations from the Crustal Movement Observation Network of China (CMONOC), we constructed an ionospheric electron density nowcast model. Through the analysis, we found that the electron density of a grid point in space has a strong correlation with its nearby grid; the farther the distance is, the smaller the correlation is, and its value tends toward 0. In order to obtain a more sparse background field covariance matrix, it was necessary to set the correlation coefficient threshold. Generally, the correlation coefficient of the point is 0 when the correlation is less than the threshold, and the covariance is 0, as well. In previous studies, many researchers directly truncate the covariance matrix by the correlation coefficient. This may cause the truncated matrix to cease to be a non-negative-definite matrix, which makes the Kalman filter divergent. In this work, the localization function was used to truncate the covariance matrix. Using the CL to sparse the covariance matrix can ensure that the truncated matrix is still a positive-definite matrix. When the sparse covariance matrix is stored in the sparse matrix format, its memory is only about 1.5% of the original matrix, which greatly reduces the requirements for the performance of the working computer, and the single calculation time is also greatly reduced. The calculation time on a personal laptop is only a few minutes, which greatly facilitates the application of the data assimilation algorithm in practical engineering.

The data assimilation algorithm can assimilate the two-dimensional observation data to the three-dimensional background field, which helps to obtain a more accurate ionospheric electron density distribution. By comparing the IRI modeling and assimilation

results with observations, we found that the assimilation algorithm effectively improves the peak accuracy of ionospheric electron density and improves the accuracy of ionospheric critical frequency. Compared with the background field, the TEC distribution and electron density distribution, after assimilation, successfully reproduced the ionospheric details. However, the assimilation algorithm does not significantly improve the peak height of electron density, which is mostly due to the fact that ground-based GNSS observation VTEC data cannot provide the vertical structure information of the ionosphere. We propose that by adding other observations to the model, such as digisonde, GNSS occultation and other data with ionospheric vertical structure information, the peak height accuracy of electron density should effectively improve. In addition, the assimilation results deviate from the observation in some cases, which may be owing to the ionospheric structure information provided by the ionospheric background model being quite different from the real ionospheric structure. Thus, the assimilation model cannot obtain accurate ionospheric structure information, leading to a deviation of the analysis from the actual observation. Therefore, constructing an accurate ionospheric background model is also a key to data assimilation.

**Author Contributions:** Conceptualization, J.Q. and Y.L.; Data curation, C.Z. and J.Q.; Formal analysis, J.Q.; Funding acquisition, C.Z.; Investigation, J.Q. and Y.L.; Methodology, J.Q., C.Z., J.Z. and Z.Z.; Resources, Y.L.; Software, Z.Z.; Validation, J.Q.; Visualization, J.Q. and J.Z.; Writing review & editing, C.Z. and Z.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key R&D Program of China, grant No. 2018YFF01-013706, 2018YFC1503506; the National Science Foundation of China, grant No. 42074187; the foundation of the National Key Laboratory of Electromagnetic Environment, grant No. 6142403180204; the Excellent Youth Foundation of Hubei Provincial Natural Science Foundation, grant No. 2019CFA054, and the research grant from the China Research Institute of Radiowave Propagation (research on low ionosphere satellite detection).

**Data Availability Statement:** The data presented in this study are openly available in <https://zenodo.org/record/6992678#.YvpBPOhBzb0> at doi:10.5281/zenodo.6992678.

**Acknowledgments:** The authors gratefully acknowledge the Crustal Movement Observation Network of China (CMONOC) for providing the GPS data and gratefully acknowledge the Meridian Space Weather Monitoring Project for the digisonde data. The IRI Fortran code can be downloaded from <http://irimodel.org/> (accessed on 10 September 2017). The CMONOC data can be downloaded from the website (<https://zenodo.org/record/6395973> (accessed on 30 March 2022)). The digisonde data can be downloaded from the website (<https://data.meridianproject.ac.cn/> (accessed on 1 May 2020)).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Scherliess, L.; Schunk, R.W.; Sojka, J.J.; Thompson, D.C.; Zhu, L. Utah State University Global Assimilation of Ionospheric Measurements Gauss-Markov Kalman filter model of the ionosphere: Model description and validation. *J. Geophys. Res. Atmos.* **2006**, *111*, A11315.1–A11315.12. [[CrossRef](#)]
2. Yue, X.; Wan, W.; Liu, L.; Zheng, F.; Lei, J.; Zhao, B.; Xu, G.; Zhang, S.-R.; Zhu, J. Data assimilation of incoherent scatter radar observation into a one-dimensional midlatitude ionospheric model by applying ensemble Kalman filter. *Radio Sci.* **2007**, *42*, 1–20. [[CrossRef](#)]
3. Bust, G.S.; Garner, T.W.; Gaussiran, T.L. Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm. *J. Geophys. Res. Earth Surf.* **2004**, *109*, A11312. [[CrossRef](#)]
4. Hajj, G.A.; Wilson, B.D.; Wang, C.; Pi, X.; Rosen, I.G. Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter. *Radio Sci.* **2004**, *39*, 1–17. [[CrossRef](#)]
5. Schunk, R.W.; Scherliess, L.; Sojka, J.J.; Thompson, D.C.; Anderson, D.N.; Codrescu, M.; Minter, C.; Fuller-Rowell, T.J.; Heelis, R.; Hairston, M.; et al. Global Assimilation of Ionospheric Measurements (GAIM). *Radio Sci.* **2004**, *39*, RS1S02. [[CrossRef](#)]
6. Yue, X.; Schreiner, W.S.; Kuo, Y.-H.; Hunt, D.C.; Wang, W.; Solomon, S.C.; Burns, A.G.; Bilitza, D.; Liu, J.-Y.; Wan, W.; et al. Global 3-D ionospheric electron density reanalysis based on multisource data assimilation. *J. Geophys. Res. Earth Surf.* **2012**, *117*, A09325. [[CrossRef](#)]

7. Yue, X.; Schreiner, W.S.; Kuo, Y.-H.; Braun, J.J.; Lin, Y.-C.; Wan, W. Observing System Simulation Experiment Study on Imaging the Ionosphere by Assimilating Observations from Ground GNSS, LEO-Based Radio Occultation and Ocean Reflection, and Cross Link. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 3759–3773. [[CrossRef](#)]
8. Lee, I.T.; Matsuo, T.; Richmond, A.; Liu, J.Y.; Wang, W.; Lin, C.C.H.; Anderson, J.L.; Chen, M.Q. Assimilation of FORMOSAT-3/COSMIC electron density profiles into a coupled thermosphere/ionosphere model using ensemble Kalman filtering. *J. Geophys. Res. Space Phys.* **2012**, *117*, 10318. [[CrossRef](#)]
9. Hsu, C.; Matsuo, T.; Yue, X.; Fang, T.; Fuller-Rowell, T.; Ide, K.; Liu, J. Assessment of the Impact of FORMOSAT-7/COSMIC-2 GNSS RO Observations on Midlatitude and Low-Latitude Ionosphere Specification: Observing System Simulation Experiments Using Ensemble Square Root Filter. *J. Geophys. Res. Space Phys.* **2018**, *123*, 2296–2314. [[CrossRef](#)]
10. He, J.; Yue, X.; Wang, W.; Wan, W. EnKF Ionosphere and Thermosphere Data Assimilation Algorithm Through a Sparse Matrix Method. *J. Geophys. Res. Space Phys.* **2019**, *124*, 7356–7365. [[CrossRef](#)]
11. Jee, G.; Burns, A.G.; Wang, W.; Solomon, S.; Schunk, R.W.; Scherliess, L.; Thompson, D.C.; Sojka, J.J.; Zhu, L. Duration of an ionospheric data assimilation initialization of a coupled thermosphere-ionosphere model. *Space Weather* **2007**, *5*, S01004. [[CrossRef](#)]
12. Hsu, C.-T.; Matsuo, T.; Wang, W.; Liu, J.-Y. Effects of inferring unobserved thermospheric and ionospheric state variables by using an Ensemble Kalman Filter on global ionospheric specification and forecasting. *J. Geophys. Res. Space Phys.* **2014**, *119*, 9256–9267. [[CrossRef](#)]
13. Gardner, L.C.; Schunk, R.W.; Scherliess, L.; Sojka, J.J.; Zhu, L. Global Assimilation of Ionospheric Measurements-Gauss Markov model: Improved specifications with multiple data types. *Space Weather* **2014**, *12*, 675–688. [[CrossRef](#)]
14. He, J.; Yue, X.; Le, H.; Ren, Z.; Wan, W. Evaluation on the Quasi-Realistic Ionospheric Prediction Using an Ensemble Kalman Filter Data Assimilation Algorithm. *Space Weather* **2020**, *18*, e2019SW002410. [[CrossRef](#)]
15. Qiao, J.; Liu, Y.; Fan, Z.; Tang, Q.; Li, X.; Zhang, F.; Song, Y.; He, F.; Zhou, C.; Qing, H.; et al. Ionospheric TEC data assimilation based on Gauss–Markov Kalman filter. *Adv. Space Res.* **2021**, *68*, 4189–4204. [[CrossRef](#)]
16. Galkin, I.A.; Reinisch, B.W.; Huang, X.; Bilitza, D. Assimilation of GIRO data into a real-time IRI. *Radio Sci.* **2012**, *47*, RS0L07. [[CrossRef](#)]
17. Angling, M.J.; Shaw, J.; Shukla, A.K.; Cannon, P.S. Development of an HF selection tool based on the Electron Density Assimilative Model near-real-time ionosphere. *Radio Sci.* **2009**, *44*, RS0A13. [[CrossRef](#)]
18. Yue, X.; Schreiner, W.S.; Lin, Y.-C.; Rocken, C.; Kuo, Y.-H.; Zhao, B. Data assimilation retrieval of electron density profiles from radio occultation measurements. *J. Geophys. Res. Earth Surf.* **2011**, *116*, 148–227. [[CrossRef](#)]
19. Mengist, C.K.; Kim, Y.H.; Ssessanga, N.; Kim, J. A Data Assimilated Regional Ionosphere Model Using the Total Electron Content from the Korean GPS Network. *J. Korean Phys. Soc.* **2018**, *72*, 826–834. [[CrossRef](#)]
20. Aa, E.; Huang, W.; Yu, S.; Liu, S.; Shi, L.; Gong, J.; Chen, Y.; Shen, H. A regional ionospheric TEC mapping technique over China and adjacent areas on the basis of data assimilation. *J. Geophys. Res. Space Phys.* **2015**, *120*, 5049–5061. [[CrossRef](#)]
21. Forsythe, V.V.; Azeem, I.; Crowley, G. Ionospheric Horizontal Correlation Distances: Estimation, Analysis, and Implications for Ionospheric Data Assimilation. *Radio Sci.* **2020**, *55*, e2020RS007159. [[CrossRef](#)]
22. Forsythe, V.V.; Azeem, I.; Crowley, G.; Themens, D.R. Ionospheric Vertical Correlation Distances: Estimation from ISR Data, Analysis, and Implications for Ionospheric Data Assimilation. *Radio Sci.* **2021**, *56*, e2020RS007177. [[CrossRef](#)]
23. Wang, S.; Huang, S.; Fang, H. Estimating of the Global Ionosphere Maps Using Hybrid Data Assimilation Method and Their Background Influence Analysis. *J. Geophys. Res. Space Phys.* **2020**, *125*, e2020JA028047. [[CrossRef](#)]
24. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
25. Rosen, I.G.; Wang, C.; Hajj, G.; Pi, X.; Wilson, B. An Adjoint Method Based Approach to Data Assimilation for a Distributed Parameter Model for the Ionosphere. In Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, FL, USA, 4–7 December 2001. [[CrossRef](#)]
26. Aa, E.; Liu, S.; Huang, W.; Shi, L.; Gong, J.; Chen, Y.; Shen, H.; Li, J. Regional 3-D ionospheric electron density specification on the basis of data assimilation of ground-based GNSS and radio occultation data. *Space Weather* **2016**, *14*, 433–448. [[CrossRef](#)]
27. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411. [[CrossRef](#)]
28. Anderson, J.L.; Anderson, S.L. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **1999**, *127*, 2741–2758. [[CrossRef](#)]
29. Gaspari, G.; Cohn, S.E. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* **1999**, *125*, 723–757. [[CrossRef](#)]
30. Houtekamer, P.L.; Mitchell, H.L. A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Weather Rev.* **2001**, *129*, 123–137. [[CrossRef](#)]
31. Anderson, J.L. A Local Least Squares Framework for Ensemble Filtering. *Mon. Weather Rev.* **2003**, *131*, 634–642. [[CrossRef](#)]
32. Moosavi, A.; Attia, A.; Sandu, A. A machine learning approach to adaptive covariance localization. *arXiv* **2018**, arXiv:1801.00548.
33. Furrer, R.; Genton, M.G.; Nychka, D. Covariance Tapering for Interpolation of Large Spatial Datasets. *J. Comput. Graph. Stat.* **2006**, *15*, 502–523. [[CrossRef](#)]

- 
34. Bilitza, D.; Reinisch, B. International Reference Ionosphere 2007: Improvements and new parameters. *Adv. Space Res.* **2008**, *42*, 599–609. [[CrossRef](#)]
  35. Bilitza, D.; Altadill, D.; Truhlik, V.; Shubin, V.; Galkin, I.; Reinisch, B.; Huang, X. International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions. *Space Weather* **2017**, *15*, 418–429. [[CrossRef](#)]